

U.S. Army - Baylor University Graduate Program in Healthcare Administration

Graduate Management Project:

An Evaluation of the Johns Hopkins Outpatient Surgery Center

Patient Satisfaction Survey

Submitted to:

Major Barry Evans, Ph.D.

November 24, 1999

Captain Dorothy L. Oakes, USAF, MSC, CHE

600 North Wolfe Street, Billings Administration 107

Baltimore, MD 21287-1607

410-955-5059

Report Documentation Page		
Report Date 00 Nov 1999	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle An Evaluation of the Johns Hopkins Outpatient Surgery Center Patient Satisfaction Survey		Contract Number
		Grant Number
		Program Element Number
Author(s)		Project Number
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) The Johns Hopkins Hospital 600 N. Wolfe St Admin 107 Baltimore, MD 21287-1607		Performing Organization Report Number
Sponsoring/Monitoring Agency Name(s) and Address(es)		Sponsor/Monitor's Acronym(s)
		Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes		
Abstract		
Subject Terms		
Report Classification unclassified		Classification of this page unclassified
Classification of Abstract unclassified		Limitation of Abstract UU
Number of Pages 70		

Acknowledgements

I am grateful for the assistance and guidance provided by Terence Cunningham, Vice President of Administration; Roberta Heath, Director of Outpatient Surgery and Special Procedures Units; Mollie Jenckes, Quality Management Analyst; and Dr. Laura Morlock, Professor of Health Policy and Management. Their support was instrumental in bringing this project to fruition.

Abstract

Changes in the health care market such as increased competition, managed care, and declining reimbursements provided incentive for health care organizations to seek better ways of doing business and many started embracing the quality culture. These changes in the health care market along with a shift to consumerism have forced virtually all health care organizations to face the challenge of defining and measuring quality. Customer or patient satisfaction is a widely accepted measure of quality. Numerous organizations use satisfaction surveys as the foundation for organizational improvement and the Johns Hopkins Health System is no different.

The purpose of this study was to evaluate the patient satisfaction survey instrument used in the Johns Hopkins Outpatient Surgery Center and to provide recommendations on how surveying patient satisfaction in this service area might be improved. The objectives of the study in regards to evaluation of the current survey instrument were two-fold. The first part of the evaluation assessed the psychometric properties of the instrument using statistical analyses. The second part of the evaluation critically evaluated the survey instrument relative to the literature.

The analyses of the survey indicated that the survey was adequate in terms of validity and reliability. Nonetheless, suggestions on how to improve the survey are offered. Improving the survey using scientific methodology is a resource-consuming process. Based on this and other facts and the desire to benchmark with other academic medical centers it is recommended that an outside vendor be used to provide patient satisfaction survey services to the Outpatient Surgery Center.

Table of Contents

Introduction	6
Literature Review	11
Purpose Statement	36
Methods and Procedures	36
Results	38
Discussion	43
Recommendations and Conclusions	55
Appendix	59
References	60

List of Tables

Table 1	Trends in Outpatient Surgery Volumes	10
Table 2	Descriptive Statistics	39
Table 3	“Our Employees” Survey Item Response Frequencies and Percentages	39
Table 4	“Our Unit” Survey Item Response Frequencies and Percentages	40
Table 5	Correlations Between Survey Item Responses	40
Table 6	Items Loadings (Correlations) with Primary Factor (Component)	41
Table 7	Summary of Simultaneous Multiple Regression Analysis for Variables Predicting Intent to Recommend (N = 2,334)	42
Table 8	Summary of Simultaneous Multiple Regression Analysis for Variables Predicting Overall Impression (N = 2,336)	42
Table 9	Summary of Unique Variance Contribution of Each Variable in Predicting Overall Impression (N = 2,336)	43

Introduction

The acceptance and proliferation of total quality management and process improvement that began in the early 1980's resulted in an increased sensitivity to exploring customer expectations and assessing what satisfied and even delighted customers. Manufacturing industries were the first to embrace the total quality movement. The seminal event that provided the impetus for organizations in the United States to look to total quality management occurred on June 24, 1980 with the airing of the television documentary If Japan Can ... Why Can't We? This highly successful documentary outlined the work of W. Edwards Deming and Joseph Juran in Japan after World War II that ultimately propelled Japan to become a leader in several manufacturing industries through quality and productivity improvements. Using statistical process control techniques and changing organizational cultures to be customer focused were the keys to Japan's turnaround after the war (Walton, 1988).

During this time of the quality revolution in the United States, the health care industry did not see an immediate need to implement quality practices and techniques. Eventually, the health care industry saw the need to espouse the tenets of managing for quality and improving processes. Many health care organizations started their quality journeys in late 1980's, as reported by the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) (1992) in their book Striving Toward Improvement: Six Hospitals in Search of Quality. This book describes the quality journey of six hospitals that had made significant progress in their quality improvement journeys. Each of the six hospitals featured began their quality journey in 1988. It seems as no coincidence that these organizations began their quality journeys one year following the National Demonstration Project (NDP). The NDP was an experiment in the application of quality management to health care. In the NDP, twenty-one health care organizations were matched up with an industrial quality management

expert from a major American company to implement quality management methods in a health care setting (Berwick, Godfrey, and Roessner, 1990).

Changes in the health care market such as increased competition, managed care, and declining reimbursements provided the incentive for health care organizations to seek better ways of doing business and many started embracing the quality culture. These changes in the health care market along with a shift to consumerism in the health care arena forced health care organizations to face the challenge of defining and measuring quality. Customer or patient satisfaction is a widely accepted measure of quality (Carey & Seibert, 1993). Support of this is clearly evident by the proliferation in the development and use of patient satisfaction instruments over the past decade (Marsh, 1999). A survey conducted in 1993 by Cooper Research found that almost 95% of hospitals surveyed used some form of patient satisfaction measurement and 83% of respondents used the patient satisfaction survey results to direct and create organizational changes (Burns, 1993).

Problem Statement

Numerous organizations, regardless of the industry, use satisfaction surveys as the foundation for organizational improvement. It is through the process of organizational improvement that organizational structure and processes are evaluated and subsequently changed to move the organization to higher levels of performance. Results from satisfaction surveys are used to indicate what service areas or processes can be improved to increase the consumer's level of satisfaction. In some organizations satisfaction survey results are used to determine rewards and even executive bonuses (Spitzer, 1988). Several organizations tie patient satisfaction to employee performance. Thirty-eight percent of organizations surveyed in a 1993 study use patient satisfaction survey results when conducting reviews of work performance (Burns, 1993).

If satisfaction surveys are to be useful to organizations they must satisfy psychometric standards, i.e. validity and reliability (Meterko, Nelson, & Rubin, 1990). If a survey provides inaccurate or unreliable results or results that are of little or no value or utility to the organization, then much time and resources are wasted by the organization in developing, preparing, and distributing the survey, and analyzing results. The organization must also consider the inconvenience to the respondent in time spent completing the survey. With so much riding on the results of satisfaction surveys it is imperative that the survey instrument be carefully constructed and adequately assessed for maximum utility.

Conditions which Prompted the Study

Increased competition in health care has led many health care organizations to seek ways to attract new consumers and maintain current consumers of health care services. In order to do this health care organizations must have adequate means to assess consumers' wants, needs, perceptions, and expectations so that they can better evaluate and improve programs and processes. In an environment of increasing demand and diminishing resources, improvement efforts must be carefully selected and appropriately focused to make the most efficient use of scarce resources.

Over the past two decades health care consumers have changed their view of health care. Many consumers are playing a more active role in their health care decisions. This is a result of several factors at work throughout the general population. First, today's health care consumers are better informed, thus reducing the asymmetry of information between patients and providers. Access to health care related information has improved dramatically with increased proliferation of internet usage. A recent survey indicated that in 1998 some 60 million Americans used the internet to search for health care related information (Ryan Report, 1999).

Another factor that has changed how consumers view health care is the backlash against managed care, which has increased awareness of health care issues. The on-going congressional debate on the “Patient Bill of Rights” is yet another factor impacting consumers’ view of health care and heightening awareness of the general public. The heightened awareness and changing view have led to more demanding patients who are increasingly critical of the quality of care and services they receive. In response to consumer demands regulatory bodies such as the Joint Commission on Accreditation of Healthcare Organizations, National Committee for Quality Assurance, and some state regulators are assessing patient satisfaction with regard to a multitude of areas. These areas include the iron triangle of cost, quality, and access as well as convenience and perception of services (Krowinski & Steiber, 1996). Additionally, some third party payers are using patient satisfaction and outcomes measures to determine with whom they will continue contracts and to whom they will award contracts (Bowers, Swan, & Koehler, 1994; Chung, Hamill, Kim, Walters, & Wilkins, 1999; Rosselli, Santalucia, Woodward, & Luecke, 1996).

In the ambulatory setting there is increased need to assess patient satisfaction due to the shift from inpatient care to outpatient care. This shift is due to technological advancements in procedures and equipment and managed care’s drive to reduce costs and length of stays. An additional push for this shift comes from consumer demand that is a result of an ideological shift in our society to have needs satisfied expediently. A research study conducted by the Health Care Advisory Board (1999a), using the Source Book of Health Insurance Data, Health Insurance Agency of America (HIAA), showed the increasing trend in outpatient surgery volumes from 1991, 1996, and projected into 2001. The total volume (in millions) of outpatient surgery procedures performed in hospital settings, freestanding surgery centers, and physician offices are as follows: 15.8 in 1991, 22.1 in 1996, and a projected 26.8 in 2001.

See Table 1 for a breakdown of volumes by procedure setting. This trend in increased outpatient surgical volumes provides further incentive to determine what satisfies patients so that efforts can be taken to improve satisfaction and develop targeted marketing to attract patients to increase market share. Additionally, the shift from inpatient to outpatient reduces the amount of contact time between health care providers and patients, further stressing the importance of assessing patient satisfaction (Noon & Davero, 1987).

Table 1. Trends in Outpatient Surgery Volumes (in millions)

	Hospital Outpatient	Freestanding Surgery	Physician Offices
1991	11.6	2.8	1.4
1996	13.8	4.9	3.4
2001	14.7	6.8	5.3

Source: Sourcebook of Health Insurance Data, HIAA, 1996; SMG Marketing Group, Inc. as compiled from research conducted by the Health Care Advisory Board.

In addition to these general market trends there are specific local conditions that prompted this study. The Johns Hopkins Medicine Senior Director of Ambulatory Operations and Practice Management questions the reliability, validity, sensitivity, and specificity of the current Outpatient Surgery Patient Satisfaction Survey instrument. The current survey instrument is an in-house developed, eleven-item questionnaire (Appendix). Eight of the items are 5-point scaled items, ranging from 1 – below expectations to 5 – exceeded expectations. One question offers a dichotomous response choice and the final two questions are open-ended questions. The current survey has been used for over four years and is yielding somewhat undifferentiated results with little variance over time. Therefore, some staff members, such as the chair of Johns Hopkins Outpatient Center Operating Room Committee, question the value and utility of this data. Additionally, the psychometric (reliability and validity) properties of the survey instrument have not been analyzed.

Literature Review

Assessing and measuring patient satisfaction has been widely discussed in the literature with a burgeoning beginning in the 1970's. This literature review will focus on defining patient satisfaction, methods of assessing patient satisfaction, the research design, survey construction, distribution, and analysis of self-administered surveys. The axiom of what gets measured gets done (or managed) seems to apply to patient satisfaction. Clinical managers, health care administrators, and executives will find it very difficult to manage and improve processes or outcomes that are not clearly defined and subsequently measured appropriately. Before going further, it is important to first define patient satisfaction.

Patient Satisfaction

Although there is no widely accepted single definition of patient satisfaction there is agreement in the literature that patient satisfaction is multi-dimensional (Chung et al.1999; Hall & Dornan 1988a; Krowinski & Steiber, 1996; Lewis, 1994; Linder-Pelz, 1982; Spitzer, 1988; Strasser & Davis, 1991; Ware, Davies-Avery, & Stewart, 1978; Ware, Snyder, Wright, & Davies, 1983; Zyzanski, Hulka, & Cassel, 1974). Strasser's and Davis's (1991) description of patient satisfaction captures fairly succinctly the essence of what was found in the literature.

Patient satisfaction is conceptually defined as patients' value judgments and subsequent reactions to the stimuli they perceive in the health care environment just before, during, and after the course of their inpatient stay or clinic visit. These value judgments and reactions will be influenced by the dispositional characteristics of the patients and their previous life and health care experiences (p. 55).

Patient satisfaction is related to perceived quality. Patient satisfaction is the patient's judgment of the quality of care received and is an important indicator in determining the quality of health care delivery (Chung et al.1999; Donabedian, 1966 and 1988; Ford, Bach, & Fottler, 1997; Marsh, 1999; Ware, et al., 1983). Donabedian (1988) states that patient satisfaction is vital in assessing the quality of the design and administration of health care organizations. Most patients assess quality in much the same way, without scientific basis, which is counter to the scientifically based method physicians use to evaluate quality (Bowers et al., 1994; Lewis, 1994; Spitzer, 1988; Tucker & Munchus, 1998; Ware & Snyder, 1975). Patients do not distinguish between the art and technical aspects of health care (Like & Zyzanski, 1987).

As noted by Eriksen (1987), Donabedian claims that patient satisfaction demonstrates the provider's or health care organization's degree of success in measuring up to the patient's expectations and values, those matters on which the patient is the expert. Many times patient perception of satisfaction is based on what is important to the patient's psychological and physical well being and comfort (Spitzer, 1988). In order for the organization to effectively delve into realm of patient satisfaction, satisfaction must somehow be measured.

Patient satisfaction is measured for a variety of reasons as previously discussed. In addition to being related to quality outcomes and quality improvement, patient satisfaction is also related to health service economics. Patients who are dissatisfied are more likely not to comply with health care providers' directions, may not return for a scheduled follow-up, may disenroll from their health care plan, and may negatively influence others in regards to the provider or the health plan (Linn, 1975; Marsh, 1999; Rosselli et al, 1996; Rubin et al. 1993; Ware & Davies, 1983; Ware & Hays, 1988).

Methods of Surveying Patient Satisfaction

There are numerous methods of surveying patient satisfaction, which can be divided into qualitative and quantitative methodologies. Qualitative research methods are those that describe in words or in writing the care or service received and the patient's experience of that care. In essence, it is entering the patient's world and assessing health care services from the patient's perspective.

Conducting a focus group to discuss patient experiences with a small group of patients who recently received care is an example of quantitative research. Quantitative research methods measure patient satisfaction by some form of a counting measure or through the use of scaled responses, such as those used in a Likert-type questionnaire (Krowinski & Steiber, 1996; Strasser & Davis, 1991).

Frequently qualitative methods are used first to explore the aspects of the service or care rendered that appear to be related to satisfaction. Qualitative research provides the researcher with what patients find to be important determinants of patient satisfaction. Additionally, qualitative research can serve as an indicator for what questions should be asked in quantitative research to adequately assess patient satisfaction. Good quantitative research is based on qualitative research. Qualitative methods are used to ascertain the particular areas of interest, while the quantitative methods are used to measure them. Quantitative research methods determine percent of patients satisfied, satisfaction intensity, and what attributes of care are contributing or detracting from patient satisfaction. (Krowinski & Steiber, 1996).

Most of the studies reviewed for this research project used some form of quantitative research method, while many used a combination of qualitative and quantitative research methods. Ford, et al. (1997) provide a comprehensive review of qualitative and quantitative research methods used for measuring patient satisfaction. The qualitative methods discussed by Ford et al. (1997) include:

managerial observations, employee feedback programs, work teams and quality circles, and focus groups. Another qualitative method discussed by Krowinski and Steiber (1996) is the in-depth interview. It may be of interest to note that Ford et al. (1997) categorize interviews as quantitative research methods.

The quantitative methods discussed by Ford et al. (1997) include: comment cards, self-administered mailed surveys, on-site personal interviews, telephone interviews, and mystery shoppers. All these methods should be self-explanatory except for the mystery shopper. In mystery shopping a trained observer encounters the health care system as a patient and records the encounter with numerical ratings so that the ratings can be compared over time.

Ford et al. (1997) provide detailed descriptions and substantial discussion of advantages and disadvantages of each research method. Developing an organizational strategy for assessing and measuring patient satisfaction that determines the best fit between the needed information and the value and the cost of the information requires knowledge of what methods are available. Management must carefully assess the advantages and disadvantages of each of the methods under consideration in relation to organizational factors. Organizational factors include facility size, type of technology, patient's ability to judge care they receive, length of service encounter, service location (inpatient or outpatient), and degree of organizational change.

Although not discussed by Ford et al. (1997), Fitzpatrick (1991a) suggests tape recording patients' feedback during interviews or discussion groups so that the recordings can be played back to staff. This supplement to the previously mentioned methods will help sensitize staff members to the patient's point of view. Other methods discussed by Lewis (1994) include diary keeping, telephone help lines, and videotaping patient feedback. There are numerous methods or combination of methods

that can be employed to measure and assess patient satisfaction. Method selection is not simple and must take into account reasons or objectives of assessing patient satisfaction as well as the organizational setting. Once the method is selected, the researcher must determine what questions to ask, which is the beginning of survey development.

Research Design and Survey Development

Survey development is an integral part of the overall research design. There is no one single way to develop a survey or design a research project to assess patient satisfaction. However, using a process that employs a methodical, scientifically based proven approach should yield the best results. There are several excellent texts that researchers can use as references that provide comprehensive guidance in the development of survey research. One such text is Measuring Patient Satisfaction for Improved Patient Services, by Strasser and Davis (1991), which is an easy to read text based on their experiences in developing surveys with the Patient Satisfaction Measurement Team at Ohio State University. This book contains a wealth of information on developing patient satisfaction surveys.

McDowell and Newell (1996) thoroughly discuss the current status of health measurement and describe several of the leading health measurement methods. According to the authors, the intent of the text is to provide students and researchers with ideas for further studies and to provide researchers with examples of sound measurement scales already in use so as to persuade them not to invent their own unique scales. Aday (1996) describes several steps in designing and conducting surveys. Aday's (1996) steps include several feedback loops reflecting the fact that designing a survey is "a dynamic, iterative, and interactive process" (p. 17). Krowinski and Steiber (1996) provide a good roadmap for designing patient satisfaction research that essentially covers the same topics as the previously mentioned authors in a concise format. According to Krowinski and Steiber (1996) the following are

the key areas to that are essential to developing the research design: identify decision-makers, determine the objectives, define patient satisfaction, select research methodology (qualitative and/or quantitative), and develop a research plan.

According to Krowinski and Steiber (1996) after the draft of the research plan is developed the key decision-makers need to review, provide input, and approve the plan. After decision-maker input and approval is received, which may mean multiple meetings and edits to the plan, the final plan should be a step-by-step guide of the research process. The decision-makers play a pivotal role in the research development and use of the research findings. Decision-makers will be more apt to use the resultant information if they believe the data were gathered using sound research methodology that they helped to develop and approve.

Constructing the Survey

After developing the research plan and progressing through it to the point that qualitative research has been conducted to determine what is creating satisfaction, it is time to develop the survey instrument. The determinants of satisfaction provide good indicators for what questions should be asked when developing quantitative surveys (Aday, 1996; Krowinski & Steiber, 1996). In addition to focusing on what patients deem important, Aday (1996) and Hall and Dornan (1988b) suggest reviewing other satisfaction surveys when designing or revising a survey.

Reviews of previous patient satisfaction survey research can provide ideas for the major concepts the researchers want to operationalize and can assist the researchers in constructing the survey. The survey construction process that Krowinski and Steiber (1996) suggest is to start with a draft of various forms of questions, perhaps even asking the same question different ways. Then group the questions to create a thematic flow. Strasser and Davis (1991) recommend grouping questions

based on the facets of satisfaction such as physician care, nursing care, and satisfaction with the facilities. But before questions can be grouped they must be developed.

The two types of questions used in developing survey items are qualitative and quantitative questions. Qualitative questions are open-ended questions, while quantitative questions are closed-ended questions that have predetermined answers and are structured numerically so that they easily lend themselves to statistical analysis. According to Krowinski and Steiber (1996) and Aday (1996) it is critical to use laymen's terms, using language that the patient will understand and is based on the patient's perspective. Additionally Aday (1996) states that the wording must capture or convey the objective (concept or construct) of the survey. Fielding, Hedley, Cheang, and Lee (1997) encourage researchers to ask patients to decide on the wording of questions and to get input from patient advisory councils and focus groups. Burroughs, Davies, Cira, and Dundgan (1999) successfully used focus groups, patient interviews, literature search, reviews of previous surveys, and consultation with subject matter experts to develop their survey.

Developing survey questions is as much an art as it is a science. The art challenges researchers to be creative and ingenious so that the questions asked speak to the respondent in the way intended. The art must be supported by the science aspect of developing and testing hypotheses to determine if the questions asked elicit accurate responses. It is through the science aspect that researchers learn about sources of errors in survey questions. One common error that researchers must be aware of is the respondent's propensity to give socially acceptable responses when asked threatening or sensitive questions (Aday, 1996).

Krowinski and Steiber (1996) offer the following advice to avoid common errors. Strasser and Davis (1991) and Aday (1996) support many of these points. The first point of advice is to avoid

double negative questions as they often create confusion, which can irritate respondents and discourage completion of the questionnaire, and can lead to misinterpretation of the question. Do not ask double-barreled questions. Double-barreled questions are those questions that are asking two unrelated questions, for example, “The physician was friendly and competent.” Avoid the use of jargon, abbreviations, and irrelevant questions. Use unambiguous words that are easy to understand. Avoid asking a global question about a particular service area. These types of question do not answer “why” and therefore provide information of little utility. Ware, Snyder, and Wright (1976) found that even when dissatisfied, respondents provide disproportionately high ratings to global questions. Do not ask questions that are too personal or objectionable. Finally, eschew the use of biased questions that may lead the respondents to a particular answer (Krowinski & Steiber, 1996).

The effect of an acquiescent response set is one other key issue discussed in the literature that researchers must consider when developing survey questions. Ware (1978) defines the acquiescent response set as the tendency of respondents to agree with opinion statements regardless of the content, i.e. a tendency to rate positive items positively and negative items negatively. Ware (1978) found an acquiescent response set in 40% to 60% of the respondents in three independent studies. Greater acquiescent response set bias was found from respondents with lower education or income levels. The development of multi-item scales, where several items combine into one score, that contain balanced response scales of positively and negatively worded items will reduce the effects of an acquiescent response set (Aday, 1996; Baker, 1991; Ware, 1978; Ware, Davies-Avery, & Stewart, 1978, Ware et al., 1983). Now that the general guidelines for developing survey questions have been discussed, the next two sections will focus on the two types of questions that can be asked on surveys.

Qualitative Questions.

Open-ended or qualitative questions are questions that give the respondents an opportunity to further develop their point of view in their own words. Qualitative questions are more effective in identifying problem areas and developing solutions (HCAB, 1999b). Qualitative questions result in more negative replies than quantitative questions and tend to generate information that is very specific, useful, and valid and thus can be more useful than quantitative data (Strasser & Davis, 1991). These advantages of qualitative questions can also be disadvantages. The variability in the length and amount of detail of the response require careful coding and complicate the analysis thus increasing the amount of time required to analyze results. The richness of the data is diluted by the amount of time required to analyze and summarize the results. Further complicating this problem is the issue of inter-coder reliability, which must be evaluated and corrected if results are to be coded accurately and reliably (Krowinski & Steiber, 1996).

Qualitative questions are relatively simple to construct, especially when the researcher keeps in mind the objectives of the research and asks only those questions that will help in meeting those objectives. The focus of qualitative questions is further exploration into the respondent's world to get to the answer of "why." Simplicity and clarity are paramount.

Quantitative Questions.

Closed-ended or quantitative questions help clarify the meaning of the question by giving the respondent a limited number of response choices from which to select. The limited responses and use of numerical scales make it easy to analyze the resultant data. However, as previously discussed quantitative questions are subject to bias and do not capture the entire scope of the respondent's perspective. Some disadvantages include that it takes a considerable amount of time to construct valid

and reliable questions and respondents may not really think about their responses. Additionally, quantitative questions may force respondents to select an opinion when they do not have one (Aday, 1996; Krowinski & Steiber, 1996; Strasser & Davis, 1991).

Before writing quantitative questions Krowinski and Steiber (1996) recommend that the researchers determine the type of information they want to garner from the respondents. According to Krowinski and Steiber (1996) there are three types of information that can be obtained from respondents: reports of facts, ratings or evaluations of providers and service setting, and reports of intended behavior. Most satisfaction questionnaires are comprised of questions regarding facts and evaluations (Krowinski & Steiber, 1996). According to Ware et al. (1976) evaluative or rating type questions are superior measures of patient satisfaction.

The response rating scale is an integral part of the closed-ended survey item or question. Typically, researchers have respondents rate the intensity or strength of their feelings along a graded continuum or rating scale (Krowinski & Steiber, 1996). There are numerous options in selecting response scales such as dichotomous (yes or no), trichotomous (yes, no, or maybe), or scales that ask respondents to rate services on a scale from 1 to 5 or 1 to 100.

Strasser and Davis (1991) recommend avoiding the use of dichotomous or trichotomous scales and advocate the use of five- or seven-point scales. The variance in response sets may be constrained when using fewer intervals, such as a dichotomous scale. Lewis (1994) indicates that sensitivity is decreased when dichotomous scales are used. The use of more intervals, such as a ten- or one hundred-point scale, may force respondents to make minute differential judgements unnecessarily. Strasser and Davis (1991) and Krowinski and Steiber (1996) recommend including a neutral midpoint so that those respondents who do not have an opinion are not forced to provide one. The exclusion of

a midpoint requires a forced choice for those respondents who are truly in the middle with no judgment either way. The studies of Meterko et al. (1990) Ware et al. (1983), Ware and Hays (1988), and Zyzanski et al. (1974) empirically support the use of a five-point scale and recommend its use in future studies.

Krowinski and Steiber (1996) discuss several types of response rating scales, satisfaction scales, quality scales, frequency scales, graphic scales, Likert-type scales, and summated scales. Satisfaction, quality, Likert-type, and summated scales were the types of scales most frequently encountered in the literature on patient satisfaction measurement. Satisfaction scales are those scales that ask respondents how satisfied they are based on a scale ranging from very satisfied to very dissatisfied. Quality scales ask respondents to rate service dimensions based on a scale that ranges from poor to excellent (Krowinski & Steiber, 1996). Aday (1996) notes that research suggests that the quality scale is superior to the satisfaction scale because it has a tendency for results to be less skewed to higher levels of satisfaction, there is a greater variability of results, and is more reflective of patient intentions to seek care.

Traditional Likert-type scales begin with a statement about the service or care rendered vice a question and the respondent is asked to determine degree of agreement or disagreement with the statement (Cooper & Schindler, 1998). Frequently the following five-point scale is used, 1 – strongly disagree, 2 – disagree, 3 – not sure, 4 – agree, 5 – strongly agree (Krowinski & Steiber, 1996). Summated scales are scales that are created by adding together the rating scores of internally consistent items. This type of scale is considered to be superior to assessing individual items, as individual item scores have a tendency to produce measurement errors (Krowinski & Steiber, 1996). Ware et al.

(1976) found inconsistent and unreliable results based on individual items when respondents were re-administered the same survey.

Krowinski and Steiber (1996) recommend using Likert-type summated scales and concur with the reasons provided by Ware et al. (1983). First, use of a uniform response scale facilitates survey completion. Second, the standardized response scales make the task of formatting the questionnaire easier. The item stem is to the left with the response scale to the right of each item requires less space on a printed page than other formats. Third, item stem statements can be easily reworded to reduce skewness. Skewness is a measure of a distribution's symmetry. Skewness of patient satisfaction survey data usually shows that the majority of responses are grouped at the positive end of the response scale. The use of multi-items scales overcomes the effect of an acquiescent response set (Lewis, 1994; Ware, Davies-Avery, & Stewart, 1978). Summated scores improve reliability. Improved reliability leads to improved confidence when scores change. Improved reliability means that the changes are a result of changes in services or patient perception vice an error in measurement (Krowinski & Steiber, 1996).

Developing qualitative and quantitative questions to adequately assess patient satisfaction is no simple task. The use of qualitative and quantitative questions on the same questionnaire can improve the measurement of patient satisfaction. The quantitative questions will provide a measure and the qualitative questions can derive more specific information. Strasser and Davis (1991) found qualitative data to offer the most useful information for health care administrators and recommend including a qualitative question within major survey sections. Asking for additional general comments will better capture more of the patient's total experience than quantitative questions allow. It may also be useful to focus the respondent's comments by asking for recommendations on how services can be improved or

what aspects of their visit they liked best (Aday, 1996; Strasser & Davis, 1991). To the contrary, Krowinski and Steiber (1996) argue that when qualitative and quantitative questions are used on the same survey instrument there is a danger that “neither will be done well” (p.132). After questions are developed researchers must focus on the layout of the survey and accompanying instructions.

Survey Layout, Introduction, and Instructions.

Although the questions are the heart of the survey instrument, careful attention to detail must be paid to the layout of the survey, the introduction, and the instructions. Appearance is crucial to obtaining the potential respondent’s participation. The printing of the survey should be of high quality (Krowinski & Steiber, 1996). The layout of the survey should be easy to follow, attractive, and easy to read and understand. The instructions must be straightforward and simple and include what the patient is to do with survey after completion (Fitzpatrick, 1991b; Krowinski & Steiber, 1996; Lewis, 1994).

According to Strasser and Davis (1991), there are multiple goals to be achieved in the introduction of the survey. First, respondent confidentiality and anonymity must be assured. Second, honest responses must be encouraged. Third, respondents must be encouraged to complete and return the survey. Finally, the introduction should be developed with a marketing mindset. Additionally, Strasser and Davis (1991) and Marks (1982) state that the introductory cover letter or paragraph is crucial. They also recommend that the chief executive officer or someone in a respected position sign the introduction or cover letter. Marks (1982) also suggests explaining the benefits to the respondent, even if they are indirect, the introduction must convince the potential respondent that completing the survey has value and is worth the effort. Lewis (1994) and Fitzpatrick (1991b) concur that it is paramount to clearly, yet simply state the purpose and use of the questionnaire as well as why the person was selected.

Once the questionnaire is prepared it must be pretested, revised, and pretested again until the respondents are able to clearly comprehend the questions, response choices, and instructions (Krowinski & Steiber, 1996). Carey and Seibert (1993) recommend revising questions based on input from focus groups and staff members. After the revision process is completed, the researcher needs to examine the survey's psychometric properties by pilot testing the questionnaire on a sample of respondents that is large enough to allow for multivariate statistical analysis (Fitzpatrick, 1991a; Krowinski & Steiber, 1996).

Sampling and Survey Distribution

In order for a survey to be of any real value the sample surveyed must be representative of the total population from which the sample was drawn. Having a representative sample is essential to being able to make generalizations from the sample to the population. Samples must be scientifically representative and of sufficient size to produce statistically significant results. Representativeness is crucial especially if the results of the survey will be used for decision-making or for determining and calculating staff rewards or bonuses (Krowinski & Steiber, 1996).

Sampling determines what members of the population will be surveyed and makes data gathering more manageable and affordable. Gathering information on an entire population is typically beyond the scope and resources of most studies (Aday, 1996). Kerlinger (1986) indicates that larger samples lead to smaller errors and smaller samples render larger errors. Additionally, larger samples provide more accurate statistics. Kerlinger's (1986) idea is not that more is better, but that the probability of selecting a deviant sample is greater in smaller samples. Meterko et al. (1990) recommend larger samples if the sample is to be divided into subgroups to provide a more focused analysis.

Random sampling is a sampling method that ensures that all members of a population have an equal chance or probability of being selected (Kerlinger, 1986). Random samples are unbiased in that they safeguard representativeness by ensuring that no population member has a better of chance of being selected and eliminates bias created by selection preference of one individual type over another (Kerlinger, 1986; Krowinski & Steiber, 1996).

According to Aday (1996), Krowinski and Steiber (1996), and Marks (1982) there are four primary probability sample designs typically used in conducting surveys. They are the simple random sample, systematic random sample, stratified sample, and cluster sample. The uncomplicated sample designs are the simple and systematic random samples while the more complicated designs that are used in complex multi-stage surveys are the stratified and cluster sample methods. Sampling design selection depends on the objectives of the survey.

Krowinski and Steiber (1996) warn against nonrandom sampling as this type of sampling brings with it significant risk that the sample generated will not be representative of the population. Examples of nonrandom samples include taking a sample that is convenient at the time of the study or creating a sample that consists of patients from one nursing unit in the hospital. These types of samples, though easy to gather data from, will render any generalizations to the population virtually useless.

According to Krowinski and Steiber (1996) there is no ideal sample size. Sample representation of the population is more important than the sheer size of a sample (Cooper & Schindler, 1998). Many factors must be considered when determining sample size. The goals of research, the level of analysis, the effect of the sample design, the expected response rate, expected proportion of eligibles, and cost will all impact determination of the sample size (Aday, 1996; Cooper & Schindler,

1998; Krowinski & Steiber, 1996; Strasser & Davis, 1991). For more detailed information on sample size see Aday (1996).

Survey distribution and frequencies of distribution depend on the goals or objectives of the research and the associated costs. Strasser and Davis (1991) do not recommend distributing surveys during the patient's stay as the patient's alertness and level of pain may impair judgment. Strasser and Davis (1991) found that distributing surveys at the time the patient is departing the facility was less than successful in an outpatient setting. The problems with on-site hand-delivered surveys at the time of departure or discharge as discussed by Strasser and Davis (1991) are as follows. First, it is easy for the survey to get lost in all the paper work and medical supplies that may be obtained during a visit. Second, patients have more important things on their mind and may discard the survey either intentionally or unintentionally. Third, the staff may be too busy to hand out the survey. The last problem is that distribution is not systematically controlled, which could result in a biased, unrepresentative sample.

Mailed surveys have advantages over surveys given during and at the end of a patient visit. First, demographic information can be gathered from the patient's medical record that can be encoded with the survey. Second, it provides the patient an opportunity to recuperate and reflect on their visit. Third, mailed surveys can be completed in the privacy of the respondent's home affording privacy and psychological distance from the health care facility. The disadvantages of mailed surveys include the cost of mailing and the delay in getting the survey to the patient. Based on the advantages, Strasser and Davis (1991) recommend the use of a mailed survey.

The final issue to be addressed in the discussion of survey distribution is timing. The fundamental question is how long after the visit should the patient be sent a satisfaction questionnaire?

According to Strasser and Davis (1991) a perfect system for measuring patient satisfaction would query patients three to six days after discharge. Zimmerman, Zimmerman, and Lund (1996) claim that Press, Ganey Associates, Inc., a well-known patient satisfaction survey firm, recommends mailing surveys to patients one to three days post discharge. This falls in line with Strasser and Davis' (1991) recommendation based on the assumption of a postal system that would deliver the survey in two days time.

Response Rates

Response rates vary from survey to survey. Interview and telephone surveys typically enjoy higher response rates, while mailed self-administered surveys have a wide range of response rates (Rubin 1990). Those studies with higher response rates usually indicate utilization of aggressive, systematic follow up. Krowinski and Steiber (1996) indicate that survey response rates can be discouragingly low, "typically in the range of 8 to 15 percent" (p. 74). However, other researchers such as Meterko et al. (1990) report high response rates to mailed self-administered surveys in the range of 60% to 70% and attribute this to tenacious follow-up methodology. The methodology used by Meterko et al. (1990) included a reminder, a second questionnaire, and finally a follow-up telephone call. Zapka et al. (1995) used the same type of follow-up methodology and realized a response rate of 63%.

Rubin's (1990) review of 36 studies reported response rates that ranged from 37% to 92%. Carey and Seibert's (1993) study of inpatient surveys conducted at 196 hospitals had rates that ranged from 40% to 97% with a 44% response rate overall. The few hospitals with the highest response rates had made follow-up telephone calls to encourage respondent participation. The Health Care Advisory Board's (1999b) research stated that hospitals reported response rates ranging from 30% to 45% and

that industry experts recommend a 40% response rate to ensure a representative sample. Another study conducted by the Health Care Advisory Board (1997) stated that profiled hospitals reported response rates of less than 50% for mailed surveys, and response rates greater than 75% for telephone interviews.

There is great variability in response rates to ambulatory surgery satisfaction surveys with reported response rates ranging from 6% to 77%. One of the hospitals profiled by the Health Care Advisory Board (1997) reported a 6% response for those who had same-day surgery where patients were given the survey on the same day of their care. Seibert et al. (1999) reported a 46% response to an outpatient surgery patient satisfaction survey that did not employ the use of follow-up activities. Carey and Seibert's (1993) study also included an ambulatory surgery survey with 125 hospitals participating achieving an unadjusted response rate of 54%. Noon and Davero's (1987) ambulatory patient satisfaction survey achieved a response rate of 77%. With the wide variability in reported response rates, researchers must employ measures to ensure higher response rates.

Numerous suggestions are given in the literature to improve response rates. Researchers should ensure a perception of importance is conveyed to the respondent (Heberlein & Baumgartner, 1978; Marks 1982; Strasser & Davis, 1991). This message can be conveyed through a personalized cover letter. Addressing the following issues in the cover letter will also help to increase response rates: guarantee confidentiality and anonymity, inform respondents that their opinions matter and that their surveys will be read, and encourage honest feedback and completion of the survey. The survey should be visually appealing, easy to read, and user-friendly (Aday, 1996; Heberlein & Baumgartner, 1978; Marks 1982; Strasser & Davis, 1991).

A self-addressed, postage paid envelope should be included with the survey. Sending follow-up postcards or letters to the people who received the questionnaires is also another method to improve response rates. This method can also serve as “thank you” for those who have already completed the survey and will serve as a re-enforcement that the respondent’s feedback is important (Aday, 1996; Heberlein & Baumgartner, 1978; Marks 1982; Strasser & Davis, 1991). Strasser and Davis (1991) recommend keeping the survey as short as possible, while Heberlein and Baumgartner (1978) found that a longer survey may send a message of increased importance.

Providing an incentive is another method employed by some researchers to increase response rates. Meterko et al. (1990) reported success using a pen sent along with the questionnaire as an incentive for respondents to complete the survey. They found the use of the pen as an incentive increased response rates by ten percentage points and was more effective than the incentive of a one-dollar bill. Aday (1996) stated that there is a significant amount of research, which indicates that monetary incentives have been shown to increase response rates without biasing results. However, she does warn that some studies indicate that incentives set unaffordable precedents. Strasser and Davis (1991) caution against the use of incentives to increase response rates as it may lead to careless completion of the survey in addition to appearing unprofessional to some respondents.

Survey Analysis and Presentation

After the survey is distributed the results must be analyzed in some meaningful manner. This discussion of survey analysis will focus on the analysis of closed-ended survey questions. The data from the completed surveys must first be prepared for analysis. The data must be coded by translating responses into numerical codes suitable for statistical analysis. For a more detailed description of developing coding rules, coding, and preparing a coding book see Aday (1996) or Strasser and Davis

(1991). The method of data entry into a computer software package or database must ensure that there are virtually no transcription errors. After the data is entered it must be analyzed.

The first step in data analysis is to clean the data. Strasser and Davis (1991) recommend that the researcher should assume that the data are wrong, until proven otherwise. Cleaning the data by conducting range and logic checks is the first step in the analysis process. This is accomplished by studying the distribution of the data. A range check is ensuring that all values fall within the assigned range. For example, if a five-point Likert-type scale was used, all data would be in the range between one and five; therefore, negative values, values of zero, or any number greater than five would indicate erroneous data. The logic check is simply ensuring that the data make sense based on the researchers knowledge of the variables. All questionable data should be verified with the original survey (Aday, 1996; Strasser & Davis, 1991).

After the data are cleaned, the next step is to run the statistical analysis. Descriptive and inferential statistics are the two types of statistics that can be used to analyze quantitative data. Descriptive statistics or univariate statistics are used to describe the sample and include such measures as frequencies, central tendency, dispersion, and shape. (Aday, 1996; Cooper & Schindler, 1998; Kerlinger, 1986; Strasser & Davis, 1991).

Inferential statistics have two basic characteristics as discussed by Kerlinger (1986) and Cooper and Schindler (1998). First, inferential statistics are those statistical methods used to make inferences of estimations about a population from a sample. Second, they are used to test hypotheses. Bivariate and multivariate statistics permit the researcher to conduct hypothesis testing that determine the existence or strength of the relationship between two or more variables. Bivariate statistical analysis, as the name implies, permits determination of the existence and/or strength of the relationship between two

variables, while multivariate statistical analysis is between two or more variables (Aday, 1996; Cooper & Schindler, 1998; Kerlinger, 1986). There are numerous types of bivariate and multivariate statistical analyses, the type of data (nominal, ordinal, interval, or ratio) and the relationship of the variables as implied by the study's objectives and hypotheses will determine the which statistical method of analysis to use (Aday, 1996).

Although there are numerous statistical analyses that could be discussed at this point such as analysis of variance, t-tests, multiple linear regression analysis, etc., only those that frequently appeared in the literature reviewed for this study will be discussed. These techniques include measures of validity and reliability, correlations, factor analysis, and regression analysis. Although a thorough discussion of each technique is beyond the scope of this project, the review of these measures will include a discussion of the definition and methodologies used.

Reliability is the degree that a measure provides consistent, stable, dependable, and predictable results (Cooper & Schindler, 1998; Kerlinger, 1986). Reliability contributes to validity "and is a necessary but not sufficient condition for validity" (Cooper & Schindler, 1998, p. 171). Kerlinger (1986) describes reliability as the "accuracy or precision of a measuring instrument" (p. 405). Kerlinger (1986) relates measurement errors and reliability; the extent of instrument unreliability determines the extent of measurement errors, i.e. the less error, the better the reliability and the more the error, the worse the reliability.

The most frequent measure of reliability used in assessing patient satisfaction survey results found during the literature review for this study was internal consistency. Split-half, test-retest, and Kuder-Richardson techniques were occasionally used, but Cronbach's alpha was the most common. All of these techniques employ the use of correlation methodology. Internal consistency methods were

used in the following studies: Abramowitz, Cote, and Berry, 1987; Baker, 1991; Burroughs et al., 1999; Carey and Seibert, 1993; DiMatteo and Hays, 1980; Forbes and Brown, 1993; Hall, Milburn, and Epstein, 1993; Hinshaw and Atwood, 1982; LaMonica, Oberst, Madea, and Wolf, 1986; Mangelsdorff, 1979; Marsh, 1999; Nelson, Hays, Larson, and Batalden, 1989; Seibert et al. 1999, Stump, Dexter, Tierney, and Wolinsky, 1995; Ware and Hays, 1988; Ware and Snyder, 1975; Ware, Wright, Snyder, and Chu, 1975; and Zyzanski et al. 1975.

Validity is multifaceted and assessing validity is controversial especially when there is no standard against which to evaluate it (Fitzpatrick, 1991a; Kerlinger, 1986; Rubin 1990; Ware et al., 1983). Validity is also the degree to which different results found with a measuring instrument reflect the actual differences of the respondents under study. External and internal validity are the two major forms of validity. External validity addresses the generalizability of the study's findings to the population, across other entities, places, and times. Internal validity is measuring what is intended to be measured (Cooper & Schindler, 1998).

There are three major types of internal validity, content, criterion-related, and construct (Baker, 1991; Cooper & Schindler, 1998; Kerlinger, 1986). The type of validity most frequently assessed in the literature reviewed was construct validity. Construct validity evaluation involves consideration of theory and the measurement instrument. How well does the instrument represent or measure the constructs under study? According to Cooper and Schindler (1998) the following methods can be used to examine construct validity: judgmental, correlation of a proposed test with an established one, convergent-discriminant correlation techniques, factor analysis, and multitrait-multimethod analysis.

In the literature correlations were used to support both reliability and validity assessments and also serve as the basis for other methods such as factor analysis. Correlations indicate the magnitude

and direction of the relationship between two variables. There are numerous types of correlations in addition to the ones previously mentioned, such as Pearson's product moment correlations, Spearman, point-biserial, etc. The following studies used some form of correlation in the analysis of survey results: Forbes and Brown, 1993; Hinshaw and Atwood, 1982; Burroughs et al., 1999; DiMatteo and Hays, 1980; Like and Zyzanski, 1987; Mangelsdorff, 1979; Nelson et al., 1989; Stein, Fleishman, Mor, and Dresser, 1993; Tucker, 1998; Tucker and Munchus, 1998; Ware and Hays, 1988; Ware and Snyder, 1975; and Ware et al. 1975. In the studies that discussed the use of correlation, survey items were most often correlated with an overall measure of satisfaction, intent to return, intent to recommend, or with each other.

As previously mentioned, factor analysis is used to examine construct validity. Nunnally (1978) states that factor analysis is a broad grouping of approaches used to conceptualize categories or groupings of variables. Factor analysis is also a broader grouping of procedures used to mathematically determine which variable, or survey item in the case of satisfaction surveys, belongs in which category or group. The following studies used factor analysis: Ambramowitz et al., 1987; Baker, 1991; Burroughs et al., 1999; Carey and Seiber, 1993; LaMonica et al., 1986; Mangelsdorff 1979 and 1994; Marsh, 1999; Stein, et al., 1993; Stump et al., 1995; McHorney, Ware, and Raczek, 1993; Seibert, et al., 1999; Ware and Snyder, 1975; Ware et al., 1975; and Ware et al., 1983.

Krowinski and Steiber (1996) describe factor analysis as a multivariate statistical technique that permits the researcher to collectively take individual responses and encapsulate them into larger patterns. Factor analysis takes individual responses and groups them into categories or factors that behave in the same manner. This grouping allows the researcher to establish key dimensions of the survey and to generate a smaller number of categories or factors to explain the variance. Hence, more

of the variance of the patient's satisfaction of health care experience is explained by fewer factors (Bryant & Yarnold, 1995; Kerlinger, 1986; Krowinski & Steiber, 1996). Jolliffe (1995) states that the core idea or central theme of factor analysis is that the measured variables represent underlying unobservable variables or factors. These underlying factors are smaller in number than the measured variables. According to Kerlinger (1986) scientific parsimony is served through factor analysis by simplifying research through identifying test measures that belong together in addition to identifying underlying dimensions.

Another method used to analyze patient satisfaction survey results is regression analysis. Regression analysis is used when the researcher is testing a hypothesis that one or more independent variables (commonly referred to as "X") influence or have a relationship to a dependent variable (commonly referred to as "Y"). The observed independent value or values are used to estimate or predict the corresponding dependent value. The changes in the dependent variable depend on and vary according to changes in the independent variable (Cooper & Schindler, 1998; Kerlinger, 1986). Some form of regression analysis was used in the following studies: Carey and Seibert, 1993; Chung et al., 1999; Davies, Ware, Brook, Peterson, and Newhouse, 1986; Hall, Milburn, and Epstein, 1993; Like and Zyzanski, 1987; Mangelsdorff, 1994; Rubin et al., 1993; Seibert, et al., 1999; Tucker, 1998; Tucker and Munchus, 1998; Ware et al., 1975.

Regression analysis helps the researcher better understand the association between variables. There are several types of regression analyses. Simple regression is where one independent variable is hypothesized to predict a single dependent variable. Multiple regression analysis is used when two or more independent variables are used to analyze the common and separate influences of the independent variables on the dependent variable (Cooper & Schindler, 1998; Kerlinger, 1986). There are other

types of regression analyses such as ordinal, logistic, and stepwise regression. The type of analysis used is determined by the objectives of the study and the type of data to be analyzed.

There are numerous techniques that can be used to evaluate and analyze the results of a patient satisfaction survey. A few researchers used only descriptive statistics to analyze their data. However, the majority of researchers in the reviewed literature used at least one or more of the more involved analytical techniques. According to Aday (1996) the key to data analysis is developing the plan of analysis prior to sitting down to analyze the data.

Measurement tools provide insight into what customers deem important. However, these measures are meaningless unless the organizational culture and executive leadership are dedicated to service excellence in meeting customer needs and exceeding their expectations (Krowinski & Steiber, 1996). A systematic method of presenting the analyzed data to administrative managers and clinicians must be developed. Results should be widely distributed so that all members of the staff can understand what patients deem important. Results of patient satisfaction surveys should be used to recognize the staff in a positive manner and should not be used for punitive purposes (Strasser & Davis, 1991).

When presenting data the researcher should use graphs that are easy to read and understand (HCAB, 1997; Strasser & Davis, 1991). Strasser and Davis (1991) recommend keeping the data presented simple, as simplicity encourages use. Although sophisticated statistical analysis may be used to in preparing the data, researchers should keep the report focused on the resultant data and associated opportunities for improvement. Most staff members will not be interested in the statistical details of the analysis. However, the researcher must be prepared to address the how and why of data analysis should the question arise.

Alternatives to Developing In-House Surveys

There are three basic alternatives to conducting homemade patient satisfaction surveys. The first option is not to conduct the survey at all, which in today's competitive, customer-focused environment is not a wise or viable alternative. The second is to use a previously developed survey, which requires gaining permission from the original survey developers. The third option is to contract with a firm that specializes in patient satisfaction surveying. A detailed discussion on selecting an outside vendor or how to find and select a previously developed survey is beyond the scope of this project. Those parties interested in more information on these topics are encouraged to review the references of this research project especially; Ware, Orgren, and Conner (1999), McDowell and Newell (1996), Hall and Dornan (1988a and 1988b), and Rubin (1990) as they will provide an overview of many of the survey instruments.

Purpose Statement

The purpose of this study was to evaluate the patient satisfaction survey instrument used in the Johns Hopkins Outpatient Surgery Center and to provide recommendations on how surveying patient satisfaction in this service area might be improved. The objectives of the study in regards to evaluation of the current survey instrument were two-fold. The first part of the evaluation assessed the psychometric properties of the instrument using statistical analyses. The second part of the evaluation critically evaluated the survey instrument relative to the literature.

Method and Procedures

The statistical examination of the survey instrument included descriptive and inferential statistics. The methods employed in the study were conducted in a systematic manner. First, all data were copied from Lotus 1-2-3 format into Microsoft Excel, a software package the researcher was familiar with, had

access to, and was certain that the Excel format could be easily transferred into the Statistical Package for the Social Sciences (SPSS). SPSS is the software package that was used to analyze the data. Additionally, this method of automated data transfer minimized the possibility of transcription errors.

The data used were results from previous surveys from fiscal years 1998 (July 1997 through June 1998) and 1999 (July 1998 through June 1999). The sample size was 2,338. The sample consisted of results from 1,097 surveys in fiscal year 1998 and 1,241 surveys in fiscal year 1999.

To obtain an overall understanding of the data, descriptive statistical analyses were performed on the data. The analyses performed included measures of central tendency, frequencies, dispersion, and shape. More sophisticated statistical analyses were performed and included bivariate and multivariate analyses. Cronbach's alpha was used to measure the survey's internal consistency or homogeneity among survey items. Cronbach's alpha is a measure of reliability and is frequently used to assess satisfaction survey reliability. Correlation coefficients were performed to understand the relationships between survey items. Factor analysis was used to assess construct validity. Regression analysis was conducted to determine the association between the dependent variable of intent to recommend and the independent variables of the eight, five-point Likert-type items. Due to the weak relationship between the variables a second regression analysis was performed using the item of overall impression as the dependent variable and the remaining seven, five-point Likert-type items as the independent variables.

Finally, a critical evaluation of the survey was conducted. The evaluation was conducted relative to the literature. The evaluation included analysis of the methodology used to develop the survey, the survey items, survey sampling and distribution, and analysis of survey results. Interviews with the individual who developed survey and who presents the survey results to staff members were

conducted to gather information for this portion of the evaluation of the survey. Additionally, informal discussions were conducted with other staff members involved in the survey process.

Results

Descriptive Statistics and Reliability

A sample of 2,338 survey results was studied. The sample consisted of results from 1,097 surveys conducted in fiscal year 1998 and 1,241 surveys conducted in fiscal year 1999. The data from the eight, five-point Likert-type items were coded using the survey results ranging from one to five. The data from the dichotomous item intent to recommend were coded one if “very likely” to recommend and zero if “not at all likely” to recommend. Table 2 displays the descriptive statistics of the survey responses. The means for each of the eight, five-point Likert-type items were skewed to the positive end of the rating scale and ranged from 4.46 to 4.79. For these same items, as the means increased the standard deviations and coefficient of variations decreased. In other words, the higher the mean, the less dispersed the data were about the mean.

Skewness and kurtosis are both measures of dispersion. Skewness is the measure of the symmetry of the distribution. All items had a negative value for skewness. The kurtosis is the measure of the peakedness or flatness of the distribution. All items had a positive kurtosis; the item with a dichotomous response scale had the highest kurtosis. The frequencies and associated percentages for the first four items on the survey under the heading of “Our Employees” are displayed in Tables 3. The frequencies and associated percentages for the last four, five-point Likert-type items under the heading of “Our Unit” are displayed in Table 4. The frequencies and associated percentages for the intent to recommend, dichotomous item were 9 (.4%) for “not at all likely” to recommend and 2,327 (99.6%) for “very likely” to recommend.

Cronbach's coefficient alpha was used to measure the internal consistency or homogeneity of the survey items. Cronbach's alpha was calculated using three different sets of variables. First, a coefficient alpha of .90 was obtained for all eight, five-point Likert-type items. A second coefficient alpha of .85 was obtained for the first four items on the survey under the heading of "Our Employees." A third coefficient alpha of .85 was obtained for the last four, five-point Likert-type items under the heading of "Our Unit." Additionally, the correlations among the five-point response scale items, discussed later, were moderate to moderately strong (.40 to .69) and statistically significant indicating further support that the survey is internally reliable.

Table 2. Descriptive Statistics

Survey Item	<u>n</u>	<u>M</u>	<u>SD</u>	Skew.	Kurt.	CV
Our Employees						
Nursing staff courtesy and respect	2,337	4.77	.48	-2.29	6.49	10.06
Staff concern for privacy	2,338	4.61	.62	-1.54	2.52	13.43
Physician courtesy and respect	2,338	4.79	.48	-2.55	7.99	10.00
Questions answered	2,338	4.69	.59	-2.19	5.91	12.65
Our Unit						
Overall impression pleasant	2,338	4.71	.55	-2.03	4.54	11.73
Pre-Op and Recovery Room Cleanliness	2,338	4.68	.56	-1.86	3.98	11.99
Acceptable noise level	2,338	4.55	.68	-1.58	2.67	15.03
Waiting area comfortable	2,338	4.46	.74	-1.39	2.09	16.57
Recommend to others	2,336	.99	.06	-16.03	255.11	6.22

Note. Data presented are fiscal years 1998 and 1999 combined. Skew = skewness, Kurt = kurtosis, and CV = coefficient of variation.

Table 3. "Our Employees" Survey Item Response Frequencies and Percentages

Response Scale	<u>Nurses</u>		<u>Privacy</u>		<u>Physicians</u>		<u>Questions</u>	
	Freq	%	Freq	%	Freq	%	Freq	%
1	2	.1	4	.2	2	.1	5	.2
2	3	.1	5	.2	5	.2	17	.7

3	44	1.9	129	5.5	47	2.0	78	3.3
4	423	18.1	634	27.1	374	16.0	507	21.7
5	1,865	79.8	1,566	67.0	1,910	81.7	1,731	74.0
Total	2,337	100.0	2,338	100.0	2,338	100.0	2,338	100.0

Note. The sum of the percentages may not add up to exactly 100 due to rounding.

Table 4. "Our Unit" Survey Item Response Frequencies and Percentages

Response Scale	<u>Impression</u>		<u>Cleanliness</u>		<u>Noise</u>		<u>Waiting Area</u>	
	Freq	%	Freq	%	Freq	%	Freq	%
1	1	.0	2	.1	5	.2	10	.4
2	13	.6	10	.4	27	1.2	29	1.2
3	71	3.0	74	3.2	145	6.2	200	8.6
4	492	21.0	553	23.7	667	28.5	747	32.0
5	1,761	75.3	1,699	72.7	1,494	63.9	1,352	57.8
Total	2,338	100.0	2,338	100.0	2,338	100.0	2,338	100.0

Note. The sum of the percentages may not add up to exactly 100 due to rounding.

Correlation Coefficients

SPSS was used to create a correlation matrix. The results of the correlation matrix are displayed in Table 5. All but one of the correlations were significant at the probability level less than .01. The correlation between concern for privacy and the intent to recommend was the weakest correlation with a probability level of .02. The correlations between the eight, five-point Likert-type items were all moderate to moderately strong with correlations ranging from .40 to .69. The weakest correlations were between each of the eight, five-point Likert-type items and the dichotomous item of intent to recommend.

Table 5. Correlations Between Survey Item Responses

Item	1	2	3	4	5	6	7	8	9
1. Nurses	--	.63**	.61**	.56**	.65**	.56**	.49**	.45**	.14**
2. Privacy		--	.56**	.53**	.57**	.58**	.55**	.49**	.05*
3. Physicians			--	.66**	.57**	.52**	.44**	.40**	.06**
4. Questions				--	.59**	.53**	.48**	.44**	.14**
5. Impression					--	.69**	.58**	.54**	.18**
6. Cleanliness						--	.64**	.55**	.11**
7. Noise Level							--	.57**	.09**
8. Waiting Area								--	.09**
9. Recommend									--

Note. ** = correlation is significant at the .01 level and * = significant at .05 level (two-tailed).

Construct Validity

Factor analysis using principal component analysis as the extraction method was used to assess construct validity. Factor analysis was performed using the eight, five-point Likert-type items. Only one component or factor emerged as a result of the principal component analysis. The one component explained 60.78% of the variance in the sample. All eight items had relatively strong correlation coefficients or loadings with the extracted factor (Table 6). Because only one factor emerged rotation of the factor was not possible.

Table 6. Items Loadings (Correlations) with Primary Factor (Component)

Items	Loading on Primary Factor
Nursing staff courteous and respectful	.80
Staff's concern for patient privacy	.79
Physician courteous and respectful	.76
Questions answered satisfactorily	.77
Overall impression pleasant	.84
Pre-Operative and Recovery Room cleanliness	.82
Noise level acceptable	.76
Family waiting area comfortable	.70

Regression Analysis

Regression analysis was performed using the dichotomous item of intent to recommend as the dependent variable and the eight, five-point Likert-type items as independent variables. The dichotomous variable was selected as the dependent variable based on the fact that intent to recommend is commonly found to be an indicator of patient satisfaction in the literature. The regression analysis of these variables resulted in the independent variables accounting for only 5.1% of the variance

in the dichotomous item of intent to recommend (Table 7). Due to the weak relationship between these variables a second regression analysis was performed using the item of overall impression as the dependent variable and the remaining seven, five-point Likert-type items as the independent variables. Overall impression was selected as the dependent variable because it is the only item that is global in nature and because it is representative of the overall construct the survey was intended to measure. The regression analysis of these variables resulted in the independent variables explaining 61.7% of the variance in the overall impression item (Table 8).

Table 7. Summary of Simultaneous Multiple Regression Analysis for Variables Predicting Intent to Recommend (N = 2,334)

Variable	<u>B</u>	<u>SE B</u>	<u>β</u>	<u>p</u>
Nursing staff courteous and respectful	.014	.004	.111	<.0005
Staff's concern for patient privacy	.011	.003	-.108	<.0005
Physician courteous and respectful	.016	.004	-.121	<.0005
Questions answered satisfactorily	.012	.003	.117	<.0005
Overall impression pleasant	.019	.004	.177	<.0005
Pre-Operative and Recovery Room cleanliness	.000	.003	.000	.993
Noise level acceptable	-.0008	.003	-.009	.768
Family waiting area comfortable	-.0003	.002	-.004	.881

Table 8. Summary of Simultaneous Multiple Regression Analysis for Variables Predicting Overall Impression (N = 2,336)

Variable	<u>B</u>	<u>SE B</u>	<u>β</u>	<u>p</u>
Nursing staff courteous and respectful	.279	.022	.242	<.0005
Staff's concern for patient privacy	.015	.017	.017	.355
Physician courteous and respectful	.085	.022	.074	<.0005
Questions answered satisfactorily	.132	.017	.142	<.0005
Pre-Operative and Recovery Room cleanliness	.302	.019	.307	<.0005
Noise level acceptable	.077	.015	.095	<.0005
Family waiting area comfortable	.081	.012	.108	<.0005

After determining the regression of the full model of the variables that predict overall impression subsequent regression analyses were performed removing one variable at a time while holding constant all other variables to determine the unique contribution of each variable. The F statistic was calculated to determine the statistical significance of the change in R^2 (Table 9).

Table 9. Summary of Unique Variance Contribution of Each Variable in Predicting Overall Impression (N = 2,336)

Variable	<u>Full</u> <u>R²</u>	<u>Reduced</u> <u>R²</u>	<u>Δ R²</u>	<u>p</u>
Nursing staff courteous and respectful	.617	.589	.028	<.01
Staff's concern for patient privacy	.617	.616	.001	<.05
Physician courteous and respectful	.617	.614	.003	<.01
Questions answered satisfactorily	.617	.607	.010	<.01
Pre-Operative and Recovery Room cleanliness	.617	.574	.043	<.01
Noise level acceptable	.617	.612	.005	<.01
Family waiting area comfortable	.617	.610	.007	<.01

Note. df 1 = 1 and df 2 = 2,329

Discussion

Statistical Analysis

All items were negatively skewed or skewed to the left (Table 2). This indicates that the distribution has a tail that stretches to the left toward the smaller values and the majority of responses are grouped in the larger value range. The positive kurtosis values indicate a leptokurtic distribution (Table 2). A leptokurtic distribution is one that is peaked with a heavy cluster of scores in the center of the curve or at either tail of the curve. The larger the kurtosis index, the more extreme the feature. To put the issue of kurtosis into perspective, a normal or mesokurtic distribution has a kurtosis index value of zero. A relatively flat or platykurtic distribution has a negative kurtosis index value (Cooper & Schindler, 1998). A review of the frequencies indicates that ninety percent or more of the results for each item were rated at a "4" or a "5" (Tables 3 and 4). The fact that the majority of the survey results were at the positive end of the rating scale was expected and is supported by the literature.

Correlations coefficients between the five-point Likert-type items were all moderate to moderately strong (Table 5). The strongest correlation at .69 was between the overall impression of the

Outpatient Surgery Center and cleanliness of the pre-operative and recovery rooms. The second strongest correlation at .66 was between physician's courteous and respectful treatment of the patient and satisfactorily answering questions before and after surgery. These first two correlations are expected, because they are related to the respondent's perception of being treated in a caring manner. The relationship between nursing staff courteous and respectful in treatment of patients and overall impression was the third strongest correlation at .65. This relationship is reasonable, and provides support that nursing staff play a large role in the patient's overall experience. The fourth strongest correlation at .64 is between cleanliness and noise level. At first blush these two areas may seem somewhat unrelated. However, both are related in that they speak to the environment and patients who are sensitive to their environment will be more keenly aware of cleanliness and sounds. Another expected relationship is the one between nursing staff's courteous and respectful treatment of patients and staff concern for privacy with a correlation of .63, which was the fifth strongest correlation. This would be expected based on the amount of time the nursing staff spends with patients and the services they provide.

All the correlations between the eight, five-point Likert-type items and the dichotomous item of intent to recommend were weak due to the fact that very few respondents intended not to recommend, the difference in the scales, and the responses given for the eight, five-point Likert-type items. There were only nine respondents who indicated that they were "not at all likely" to recommend the Johns Hopkins Outpatient Surgery Center to friends or family. Of these nine respondents six of them scored at least two of the eight items at a "4" or above. Two of these six respondents scored all eight items at a "5." One respondent rated seven items at a "5" and one item at a "4." While two other respondents scored at least half the items at a "4" or above. Fundamentally, these responses point out that although the

respondents' expectations were exceeded in the areas surveyed, they did not intend to recommend.

The results of this analysis indicate a few possibilities: respondents misinterpreted the scale, respondents misinterpreted the questions, transcription errors occurred when the data were coded into the survey result database, key determinants of patient satisfaction for these individuals are missing from the survey, or some other unexplained phenomenon occurred.

The overarching construct that was used in the development of the survey was patient perceptions or impressions of key areas. Construct validity is supported by the fact that one component emerged from the eight, five-point Likert-type items when they were factor analyzed. This one component explained a significant amount of the variance in the sample. The explanation of 60.78% of the variance by one component and the strong loadings (Table 6) indicate that the grouping of the items on the survey is appropriate.

The initial regression analysis that was conducted using the eight, five-point Likert-type items as independent variables accounted for only 5.1% of the variance in the dichotomous item intent to recommend as the dependent variable (Table 7). The small amount of variance accounted for can be explained by the reasons previously discussed in regards to the correlations between the eight, five-point Likert-type items and the dichotomous item of intent to recommend. Based on the reasons previously discussed, although some respondents found that their expectations had been exceeded by the Outpatient Surgery Center they would not recommend Johns Hopkins to their friends or families. Therefore, satisfaction or dissatisfaction with the services provided was not a clear-cut predictor of intent to recommend or not recommend.

The second regression analysis was conducted using overall impression as the dependent variable and the remaining seven, five-point Likert-type items as independent variables (Table 8). The

independent variables explained 61.7% of variance in the overall impression. This indicates that survey items adequately predict the overall impression by accounting for a significant amount of the variance and also provides further support of the construct validity. The small amount of unique variance accounted for by each variable indicates that there is a considerable amount of shared variance among the seven, five-point Likert-type items (Table 9). This also offers additional support of construct validity.

Both the full and reduced models indicate that the item of staff's concern for patient privacy has appreciably lower probabilities than the other variables. In fact, the regression analysis results of the full model show that this item is not statistically significant. This indicates the item's ability to predict a patient's overall impression could be due more to chance alone than to a true relationship between the variables when regressed simultaneously.

The statistical analysis of the survey indicates that the survey is no different than the majority of the patient satisfaction surveys in that patients tend to rate their health care experiences toward the positive end of the scale. The survey is valid and reliable. The survey measures what it was intended to measure and has high internal consistency. However, a more thorough analysis of this survey is limited due to the nature of patient satisfaction surveys and the absence of a standard against which to assess validity. Additionally, the statistical analysis of this survey is further limited by the fact that incomplete surveys were not included in the data analyzed. Those surveys are rendered useless and discarded; and are not entered into the survey result database. The inclusion of incomplete surveys may have provided different results.

Critical Evaluation of Survey Relative to the Literature

The survey (see Appendix) was developed through the efforts of one individual, the Director of Outpatient Surgery and Special Procedure Units. The goal of developing the survey was to obtain a real time assessment of patients' perceptions or impressions of key areas. Two of the major objectives in developing the survey were that the survey be short and simple. Another objective was to control costs in surveying patient satisfaction. Although no established methodology was used to develop the survey the following is an outline of the how the survey was developed.

Development of the survey involved a literature review, a review of surveys, and discussions with a small number of key staff members. The survey was developed to assess patient impressions of important measures identified in the literature and central values of the Outpatient Surgery Center. The survey items were based on survey items found on questionnaires developed by one of the better-known patient satisfaction survey firms. After the survey was developed it was immediately put into use.

The process used to develop the survey followed many of the suggestions offered in the survey development literature. Such as literature review, review of previously developed surveys, and staff input. However, survey development could have been improved through a pilot of the survey and by involving patients in the development process (Aday, 1996; Carey & Seibert, 1993; Fitzpatrick, 1991a, Krowinski & Steiber, 1996; Strasser & Davis, 1991). The involvement of patients might have included asking for their opinions and input on the questions asked and the wording of the survey items. The survey items were those areas important to the staff, but discussions with patients might have revealed other areas of importance. Clearly the survey items measure what the designer wanted measured. However, if the intent is to truly measure patient satisfaction, finding out what patients deem to be important is critical and may result in a revision of the survey.

Overall, the survey is aesthetically pleasing and professionally duplicated. The survey might be more attractive with the Johns Hopkins Outpatient Center logo located at the top of the survey. The layout of the survey is easy to follow and self-explanatory. The size of the print is adequate, but may be considered a bit small by some respondents. Also, the space provided to write responses in may be inadequate for some respondents. The objective of keeping the survey small produces challenges for the layout. Patients are verbally told what to do with the survey, i.e. place it in a survey drop box located next to each elevator lobby area or drop it in a mailbox. Although it is self-evident that the survey can be placed in a mailbox, no instructions as to what to do with the survey are provided on the survey. A short sentence of what the respondent should do with the completed survey may help improve response rates (Fitzpatrick, 1991b; Krowinski & Steiber, 1996; Lewis, 1994).

The introductory paragraph is positively worded and attempts to convey that the patient's opinion is important. However, it does not come right out and say "your opinion is important to us." The paragraph starts out with an optimistic, yet seemingly premature thank you for completing the survey. A thank you is also included at the end of the survey; saying thank you twice when space on the survey is at a premium may be unnecessary. Also, the comment that "We hope your surgical experience was pleasant" seems inappropriate. This researcher doubts that many patients would consider a surgical experience pleasant. Finally, the last sentence implants the idea that the care is high quality, which may set the stage for an acquiescent response set. A suggested introductory paragraph might read; "Your opinion is important to us and is key to improving the care and services we provide. Please take the time to tell us about your visit."

The survey item stems were similar to Likert-type item statements about different aspects of the patient's experience. All statements were positively worded which is counter to recommendations

found in the literature to balance positively and negatively worded items to reduce the effects of an acquiescent response set (Aday, 1996; Baker, 1991; Strasser & Davis, 1991; Ware, 1978; Ware et al., 1978; Ware et al., 1983). Mangelsdorff (1979) suggested using item stems that are neutrally worded. A more balanced set of survey items or neutrally worded item stems might provide more variability and less bias in results.

The majority of the statements had a single idea or related theme for each item as suggested in the literature (Aday, 1996; Krowinski & Steiber, 1996; Strasser & Davis, 1991). Although the statement “My questions before and after surgery were answered in a satisfactory manner” had a single theme, the times were different. How is a patient supposed to answer the question if questions were answered before surgery, but not afterwards? And is the statement attempting to assess the manner in which questions were answered or whether the content of the answers provided was satisfactory? A revised statement for this item might read, “My questions about my surgery were answered to my satisfaction” or “My questions about my surgery were answered in a way that I understood.”

The statement “My overall impression of the Outpatient Surgery Center was pleasant” is listed as the first item in the section “Our Unit” and should be moved to be the last item. This is a global type question and should be asked last, if at all (Strasser & Davis, 1991). More importantly, what type of information do the answers to this item provide? If the replies are negative this item provides no information to the Outpatient Surgery Center as why respondents do not have favorable impressions. This item could be deleted based on its lack of utility and respondents’ tendencies to answer global questions with disproportionately higher ratings even when dissatisfied (Krowinski & Steiber, 1996; Ware et al., 1976).

The five-point response rating scale used for eight of the survey items ranged from “exceeded expectations” to “below expectations” with “met expectations” as a mid point. This type of rating scale was not discussed in the literature. A traditional five-point Likert scale or a satisfaction scale might have been more appropriate. A five-point Likert type would have responses ranging from strongly disagree to strongly agree, while a satisfaction scale would have responses ranging from very dissatisfied to very satisfied. Additionally, a five-point Likert scale could have been used to provide increased response categories for the item regarding recommending the Johns Hopkins Outpatient Surgery Center to friends and family. The dichotomous response scale does not provide opportunity for those responses between “very likely” and “not at all likely.”

The amount of space provided for names of employees who impressed the respondent and suggestions seems inadequate and may convey to the respondents that the Outpatient Surgery Center is not very interested in what they have to say. An improved approach might be to add more space for written responses. This could be accomplished by adjusting the alignment of “Date of visit” and the optional area for “Name” and “Address.” Another suggestion would be to delete “Optional,” “Name,” and “Address” and replace it with a statement such as, “If you would like to be contacted, please provide your name and how we can reach you.” Then a line or two could be provided for the respondent to enter their phone number or address.

The nursing staff distributed the survey to patients on the day of the patient’s outpatient surgery. Theoretically, all patients were given a survey to complete. In discussions with the nursing staff it was difficult to discern if the survey was given to 100% of the patients. This method of survey distribution may lead to sampling errors and inaccuracies in calculating response rates. The nurses may either purposefully or inadvertently neglect to give every patient a survey to complete. This can lead to

sampling errors, which may bias results and bring into question the representativeness of the sample and the generalizability of the results. Another area of concern with handing out the surveys at the time of the outpatient surgery procedure is that patients may lose the survey with all the paper work and medical supplies they receive. Although this type of survey distribution may be convenient and inexpensive the disadvantages outweigh the benefits (Strasser & Davis, 1991). Mailing the self-administered surveys is offered as a suggestion to improve survey distribution. However, this would increase the cost of surveying.

Another issue related to sample representativeness and the generalizability of the results is the collection of demographic data. The survey does not include demographic variables such as age, gender, race, ethnicity, procedure code, etc. Without this information it is nearly impossible to determine if the sample is representative of the population of patients who receive care at Johns Hopkins. With this in mind, generalizations from the sample to population must be made cautiously, if at all. If changes are made based on the results of an unrepresentative sample, the changes may provide less than desirable outcomes.

Response rates for this survey over the last three fiscal years have been maintained at a consistent rate of 20%. This is not an abysmally low response rate, but leaves opportunity for improvement. The manner in which the response rates are calculated assumes that every patient who had surgery received a survey. As previously mentioned the survey distribution method does not ensure all patients receive a survey. This will artificially lower the response rate. Further confounding the low response rate is the fact that surveys that are not completely filled out are not recorded as returned surveys. These incomplete surveys are discarded and not used in the data analysis. All returned surveys should be counted when calculating the response rate. Additionally, the incomplete survey

results should also be recorded in the survey result database. The exclusion of these data could bias the results. These data could also provide insight into evaluation of the survey questions. Perhaps one question is frequently not answered.

The Outpatient Surgery Center does not employ any form of follow-up mechanism. Using some type of follow-up mechanism should result in improved response rates. A postcard reminder that doubles as a thank you for those who have responded is a relatively inexpensive way to increase response rates. A second survey and follow-up phone calls are other ways response rates could be increased (Carey & Seibert, 1993; Meterko et al., 1990; Rubin, 1990).

If a mailed self-administered survey were used, a magnet with key phone numbers of the Johns Hopkins Outpatient Center could be enclosed as an incentive and thank you for completing the survey. The magnet is only one example used as an illustration, other types of inexpensive trinkets such as an ink pens, key chains, or calendars could serve as incentives. In addition to serving as an incentive, the magnet could double as a marketing tool to provide patients easy access to important Johns Hopkins phone numbers that could be used to make future appointments. First, a decision at the organizational level must be made to approve the use of such an incentive. This type of incentive, in only one area may lead patients to expect similar incentives in other areas. The use of an incentive could start a costly precedent for the entire organization. Before the decision to use incentives is made the cost of creating the magnets and mailing the surveys would need to be weighed against the potential benefit. The Johns Hopkins Planning and Marketing and Financial Analysis Departments could be helpful in performing such an analysis.

The returned surveys are reviewed by the Center's Performance Improvement Nurse Coordinator. If a respondent rates an item at a "1" or a "2" or has negative comments and provides a

name, the Performance Improvement Nurse Coordinator attempts to contact the respondent by telephone to discuss his or her concern. Two attempts are made to contact the respondent. The surveys are forwarded to the Director of Outpatient Surgery and Special Procedure Units.

The Director reviews the surveys paying particular attention to the comments. In reviewing the comments the Director looks for trends that may indicate need for process improvement or managerial intervention. The comments are passed on to staff members, especially the positive comments. The positive comments are forwarded to staff members with a personal note from the Director. The negative comments specifically regarding physicians are forwarded directly to the physician. The Director discusses negative comments regarding the Outpatient Surgery Center staff members in general with the staff during regularly scheduled staff meetings. Comments regarding specific staff members are discussed with the individual staff member. Other comments or general trends in comments are also discussed at staff meetings. As discussed in the literature the results are not used in a punitive nature (Strasser & Davis, 1991).

The results of the survey are manually entered into a Lotus 1-2-3 spreadsheet and are analyzed by calculating the response rate and arithmetic mean for each survey item on a monthly basis. The response rates are calculated by dividing the number of returned surveys by the number of patients who had surgery during the month. The monthly response rates and means for each item are trended and compared to the previous year. Survey results are discussed with the staff and posted in the staff lounge. Additionally, up until April 1999, the results were reported to the Johns Hopkins Outpatient Center Operating Room Committee. In April, the committee chair raised concerns that the same information was reported at each meeting and suggested the survey be improved to increase sensitivity.

Patient satisfaction survey results are reviewed and discussed at the Outpatient Surgery and Special Procedure Units Performance Improvement Committee. However, there is no systematic way to track process improvement efforts back to patient satisfaction survey results. When process improvements are made as a result of patient satisfaction survey results there is no formal mechanism in place to link the two together. The Outpatient Surgery and Special Procedure Units Performance Improvement Committee minutes should provide the formal link of areas identified for improvement and process improvements. A patient satisfaction committee comprised of key staff members could be chartered to evaluate the survey results and recommend improvements that are forwarded to the Performance Improvement Committee. This committee could also follow when changes or improvements are implemented and any subsequent changes in survey results.

The survey results seem to be adequately analyzed and are used to make improvements and recognize outstanding staff members. However, a more thorough analysis of the data using the descriptive and inferential statistics discussed in the Literature Review and used in this study would provide a better evaluation of the survey and its results. A more stringent evaluation of the survey could lead to an improved survey that more accurately measures patient satisfaction. Additionally, the written comments from survey respondents could be tracked and trended to derive more specific information and to provide a better insight into the overall picture of patients' perceptions of their experiences.

Overall, the statistical analysis and the critical evaluation of the survey indicate that the survey is adequate, as it measures what it was intended to measure. However, the representativeness of the sample is questionable. The analyses also point toward opportunities for improvement.

Recommendations and Conclusions

Although the analyses of the survey profess its adequacy, the survey could be improved by implementing many of the recommendations and suggestions offered in the Discussion section of this paper. If the intent of the survey is to truly measure patient satisfaction, the focus of the survey should be changed. Instead of measuring patient perceptions of what the Director values as important, the staff should find out what the patients deem to be key determinants of satisfaction. This would lead to developing an entire a new survey.

Developing a new in-house satisfaction survey using a scientific methodology, as suggested in any one of a number of references (Aday, 1996; Krowinski & Steiber, 1996; McDowell & Newell, 1996; Strasser & Davies, 1991), is a time consuming process. A process that requires senior management support and takes the dedication of key staff members to contribute time and effort and gather patient feedback. Based on this fact and the Ambulatory Operations and Practice Management Senior Director's desire to benchmark against other academic medical centers this researcher recommends the use of an outside vendor such as Press, Ganey Associates, Inc.; Gallup; or the Picker Institute.

Over a year ago the Johns Hopkins Health System contracted with Press, Ganey Associates, Inc. to provide patient satisfaction survey services to the Hopkins family of health care business entities. Surveying of inpatient units and many outpatient clinics began in January 1999. The Outpatient Surgery Center could be added to this contract without any direct or indirect cost the department. This is an advantage over revising and administering the in-house survey in that the resources that would have been expending on the in-house survey can be used for other initiatives.

During the process of conducting this study the Outpatient Surgery Center was added to Press, Ganey Associates, Inc. contract. It is also important to note that the Outpatient Surgery Center discontinued the survey and started a pilot of a purely qualitative instrument in late October 1999. A qualitative instrument was selected because the qualitative questions from the survey analyzed in this study provided valuable information to the department. The qualitative survey will be used until the end of the calendar year. In January 2000, the Press, Ganey Associates, Inc. survey will be used to assess patient satisfaction in the Outpatient Surgery Center.

In addition to a resource savings to the department, the Press, Ganey Associates, Inc. survey offers several advantages over an in-house survey. First, Press, Ganey has several years of experience in the business of assessing patient satisfaction. As a result of their experience Press, Ganey has a database of several hundred hospitals and healthcare facilities with which they can provide comparative analyses. They also offer the flexibility of comparing with all hospitals in the database as well as a custom group selected by the client. The survey questions used by Press, Ganey are developed using scientific methods and patient input. The questions asked on the survey are tested for validity and reliability.

The Press, Ganey questionnaire is very comprehensive and covers the continuum of the patient's surgery from registration, ancillary services, pre-surgery through post-surgery. Press, Ganey provides demographic information that can be used to determine the representativeness of the survey sample to the population. They also provide sophisticated statistical data analysis. Survey results are provided in varying formats, numerical and graphical, that are relatively easy to interpret. Press, Ganey staff members provide on-site training to ensure that key staff members understanding the survey results.

In each functional unit's survey report Press, Ganey provides a Priority Index that indicates the top areas that should be focused on to improve patient satisfaction.

The biggest limitation of the Press, Ganey survey is that reports of the quantitative questions are only provided every six months. This was a decision made by the organization and was based the increased cost of receiving survey results more frequently. Although the quantitative results are only received twice a year the comments from the qualitative questions asked at the end of each section on the questionnaire are sent once a month.

Press, Ganey will begin surveying patients who have had outpatient surgery starting in January 2000. Before surveying begins the Director of Outpatient Surgery and Special Procedures Units should discuss and review the questionnaire with the staff so that they know what areas the survey will be evaluating. Also, the department should consider developing scripting for particularly sensitive areas such as privacy. When a nurse closes a curtain or covers a patient with a sheet or blanket to give the patient more privacy, the staff members should explain to the patient in simple terms what they are doing and why. For example, when the curtain is drawn, the nurse could say, "Mr. Soandso, let me close this curtain to give you more privacy." This way the patient will realize that the staff is concerned about their privacy.

In addition to preparing the staff for the upcoming surveying by Press, Ganey, the Director should also develop an action plan for addressing the results of the survey when they are received. Key staff members should be trained on the interpretation of the survey results, if they have not received the training. Additionally, the Director and other key staff members can meet with members of the Hopkins Metrics Committee, who have received special training in regards to interpreting the Press, Ganey survey results. Additionally, this committee is working closely with the Performance Improvement

Committee to ensure that the survey results lead to improved performance. As a result they may be able to offer sound recommendations on how to plan for taking action after receiving the survey results.

The first report of the quantitative survey results was provided to Johns Hopkins in September 1999. The next quantitative survey results are due out in February 2000. The Director should take advantage of any group training sessions or meetings that are offered in regards to the survey. Having a thorough understanding of the survey and the interpretation of the results will prove beneficial when the quantitative results from the Outpatient Surgery Center are published in the summer of 2000. Additionally, the Director should talk with other functional unit directors who have received survey results to see how they handled the results and to gain insight into lessons learned in dealing with survey results.

Although the Director would have preferred to be included in the initial contract with Press, Ganey Associates, Inc., the delay in being added to the contract puts the department in a unique position. The department can take advantage of lessons learned by other functional units involved in the surveying process. However, the advantage can only be realized through aggressive pursuit of lessons learned and thoughtful involvement with those currently receiving survey results. The resources to gain a full understanding of the survey and its results are available for the asking. With sound planning and careful execution of an action plan, the results of the Press, Ganey patient satisfaction survey can prove be instrumental in improving the care provided in the Outpatient Surgery Center.

Appendix

JOHNS HOPKINS OUTPATIENT SURGERY SATISFACTION SURVEY

Thank you for completing our Satisfaction Survey. We hope your surgical experience was pleasant. Please feel free to add any comments or suggestions that will help us maintain our high standard of care.

	Below Expectations		Met Expectations		Exceeded Expectations
Our Employees					
The nursing staff treated me with courtesy and respect	1	2	3	4	5
The staff was concerned about my privacy	1	2	3	4	5
My physicians treated me with courtesy and respect	1	2	3	4	5
My questions before and after surgery were answered in a satisfactory manner	1	2	3	4	5
Our Unit					
My overall impression of the Outpatient Surgery Center was pleasant	1	2	3	4	5
The Pre-Operative area and the Recovery Room were clean	1	2	3	4	5
The noise level on the unit was acceptable	1	2	3	4	5
The area where my family waited was comfortable	1	2	3	4	5

I would recommend the Johns Hopkins Outpatient Surgery Center to my friends or family.

_____ Not at all likely _____ Very likely

Name(s) of Any Employee that Impressed You _____

Suggestions on How We May Improve _____

Date of Visit _____

Optional:

Name _____

Address _____

City State Zip _____

Thank you for your comments

© 1995

References

- Aday, L. (1996). Designing and conducting health surveys: A comprehensive guide (2nd ed.). San Francisco: Jossey-Bass Publishers.
- Ambramowitz, S., Cote, A. A., & Berry, E. (1987). Analyzing patient satisfaction: A multianalytic approach. Quality Review Bulletin, 13, 122-130.
- Baker, R. (1991). The reliability and criterion validity of a measure of patient's satisfaction with their general practice. Family Practice, 8 (2), 171-177.
- Berwick, D. M., Godfrey, A. B., & Roessner, J. (1990). Curing health care. San Francisco: Jossey-Bass publishers
- Bowers, M. R., Swan, J. E., & Koehler, W. F. (1994). What attributes determine quality and satisfaction with health care delivery? Health Care Management Review, 19, 49-55.
- Burns, J. (1993). Patient satisfaction data a valuable tool for most hospital execs –survey. Modern Healthcare, 23, 54.
- Bryant, F. B. & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics. Washington, DC: American Psychological Association.
- Burroughs T. E., Davies, A.R., Cira J. C., & Dundgan, W. C. (1999). Understanding patient willingness to recommend and return: A strategy for prioritizing improvement opportunities. Joint Commission Journal on Quality Improvement, 25 (6), 217-287.
- Carey, R. G. & Seibert, J. H. (1993). A patient survey system to measure quality improvement questionnaire reliability and validity. Medical Care, 31 (9), 834-845.

Chung, K. C., Hamill, J. B., Kim, H. M., Walters, M. R., & Wilkins, E. G. (1999). Predictors of patient satisfaction in an outpatient plastic surgery clinic. Annals of Plastic Surgery, 42, 56-60.

Cooper, D. R. & Schindler, P. S. (1998). Business research methods (6th ed.). Boston MA: Irwin McGraw-Hill.

Davies, A. R., Ware, J. E. Brook, R.H., Peterson, J. R., & Newhouse, J. P. (1986). Consumer acceptance of prepaid and fee-for-service medical care. Health Services Research, 21, 429-452.

DiMatteo, M. & Hays, R. (1980). The significance of patients' perceptions of physician conduct. Journal of Community Health, 6 (1), 18-33.

Donabedian, A. (1966). Evaluating the quality of medical care. Milbank Memorial Fund Quarterly, 44 (2), 166-203.

Donabedian, A. (1988). The quality of care: How can it be assessed? JAMA, 260, 1743-1748.

Eriksen, L. R. (1987). Patient satisfaction: an indicator of quality? Nursing Management, 18, 31-35.

Fielding, R., Hedley, A., Cheang, J., & Lee, A. (1997). Methods of surveying patients' satisfaction. BMJ, 314, 227.

Fitzpatrick, R. (1991a). Surveys of patient satisfaction: I – Important general considerations. BMJ, 302, 887-889.

Fitzpatrick, R. (1991b). Surveys of patient satisfaction: II – Designing a questionnaire and conducting a survey. BMJ, 302, 1129-1132.

Forbes, M. L. & Brown, H. N. (1995). Developing an instrument for measuring patient satisfaction. AORN, 61 (4), 737-743.

Ford, R. C., Bach, S. A., & Fottler, M. D. (1997). Methods of measuring patient satisfaction in health care organizations. Health Care Management Review, 22 (2), 74-89.

Hall, J. A. & Dornan, M. C. (1990). Patient sociodemographic characteristics as predictors of satisfaction with medical care: A meta-analysis. Social Science and Medicine, 30, 811-818.

Hall, J. A. & Dornan, M. C. (1988a). Meta-analysis of satisfaction with medical care: Description of research domain and analysis of overall satisfaction. Social Science and Medicine, 27, 637-644.

Hall, J. A. & Dornan, M. C. (1988b). What patients like about their medical care and how often they are asked: A meta-analysis of the satisfaction literature. Social Science and Medicine, 27, 935-939.

Hall, J. A., Milburn, M. A., & Epstein, A. M. (1993). A casual model of health status and satisfaction with medical care. Medical Care, 31 (1), 84-94.

Health Care Advisory Board. (1997). Patient satisfaction. Fact Brief (October 1997). Washington DC: Health Care Advisory Board.

Health Care Advisory Board. (1999a). Inpatient and outpatient surgery trends: A custom report prepared from previous research. Washington DC: Health Care Advisory Board.

Health Care Advisory Board. (1999b). Patient satisfaction survey instruments. Fact Brief (April 1999). Washington DC: Health Care Advisory Board.

Heberlein, T. A. & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. American Sociological Review, 43 (4), 447-462.

Hinshaw, A. S. & Atwood, J. R. (1982). A patient satisfaction instrument: Precision by replication. Nursing Research, 31 (3), 170-175.

Joint Commission on Accreditation of Healthcare Organizations (1992). Striving toward improvement: Six hospitals in search of quality. Oakbrook, Terrace, IL: JCAHO.

Jolliffe, I. (1998). Factor analysis overview. In P. Armitage & T. Colton (Eds.), Encyclopedia of biostatistics. West Sussex, England: John Wiley & Sons Ltd.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rhinehart and Winsotn, Inc.

Krowinski, W. J. & Steiber, S. R. (1996). Measuring and managing patient satisfaction (2nd ed.). American Hospital Publishing.

LaMonica, E. L., Oberst, M. T., Madea, A. R., & Wolf, R. M. (1986). Development of a patient satisfaction scale. Research in Nursing and Health, 9, 43-50.

Lewis, J. R. (1994). Patient views on quality care in general practice: Literature review. Social Science and Medicine, 39, 655-670.

Like, R. & Zyzanski, S. J. (1987). Patient satisfaction with the clinical encounter: Social psychological determinants. Social Science and Medicine, 24, 351-357.

Linder-Pelz, S. (1982). Toward a theory of patient satisfaction. Social Science and Medicine, 16, 577-582.

- Linn, L. S. (1975). Factors associated with patient evaluations of health care. Milbank Memorial Fund Quarterly, 53, 531-548.
- Mangelsdorff, A. D. (1979). Patient satisfaction questionnaire. Medical Care, 17, 86-90.
- Mangelsdorff, A. D. (1994). Patient attitudes and utilization patterns in Army medical treatment facilities. Military Medicine, 159, 686-690.
- Marks, R. G. (1982). Designing a research project: The basics of biomedical research methodology. Belmont, CA: Lifetime Learning Publications.
- Marsh, G. W. (1999). Measuring patient satisfaction outcomes across provider disciplines. Journal of Nursing Measurement, 7 (1), 47-62.
- McDowell, I. & Newell, C. (1996). Measuring health: A guide to rating scales and questionnaires (2nd ed.). New York: Oxford University Press.
- McHorney, C. A., Ware, J. E., & Raczek, A. E. (1993). The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical Care, 31 (3), 247-263.
- Meterko, M., Nelson, E. C., & Rubin, H. R. (1990). Patient judgments of hospital quality: Report of a pilot study. Medical Care, 28 (9) (Suppl.), S1-S55.
- Nelson, E. C., Hays, R. D., Larson, C., & Batalden, P. B. (1989). The patient judgment system: reliability and validity. Quality Review Bulletin, 15 (6), 185-191.
- Noon, B. E. & Davero, C. C. (1987). Patient satisfaction in a hospital-based day surgery setting, AORN, 46, 306-312.
- Nunnally, J. C. (1978). Psychometric theory. New York: McGraw-Hill.

Rosselli, V. R., Santalucia, C. A., Woodward, A. B., & Luecke, R. W. (1996). Patient satisfaction under managed care: What every financial manager should know. Journal of Health Care Finance, 22, 61-66.

Rubin, H. R. (1990). Can patients evaluate the quality of hospital care? Medical Care Review, 47 (3), 267-326.

Rubin H. R., Gandek, B., Rogers, W. H., Kosinski, M., McHorney, C. A., & Ware, J. E. (1993). Patients' ratings of outpatient visits in different practice settings. JAMA, 270 (7), 835-840.

Ryan Report (1999). News and trends. Ryan Report Innovations, Trends & Insights, I (I), September 1999.

Seibert, J. H., Brien, J. S., Maaske, B. L., Kochurka, K., Feldt, K., Fader, L., & Race, K. E. (1999). Assessing patient satisfaction across the continuum of ambulatory care: a revalidation and validation of care-specific surveys. Journal of Ambulatory Care Management, 22 (2), 9-26.

Spitzer, R. B. (1988). Meeting consumer expectations. Nursing Administration Quarterly, 12, 31-39.

Stein, M. D., Fleishman, J., Mor, V., & Dresser, M. (1993). Factors associated with patient satisfaction among symptomatic HIV-infected persons. Medical Care, 31 (2), 182-188.

Strasser, S. & Davis, R. M. (1991). Measuring patient satisfaction for improved patient services. Ann Arbor, MI: Health Administration Press, American College of Healthcare Executives.

Stump, T. E., Dexter, P. R., Tierney, W. M., & Wolinsky, F. D. (1995). Measuring patient satisfaction with physicians among older diseased adults in a primary care municipal outpatient setting: An examination of three instruments. Medical Care, 33 (9), 958-972.

Tucker, J. L. (1998). The importance of caring as a determinant of patient satisfaction among active duty beneficiaries. Military Medicine, 163 (11), 758-760.

Tucker, J. L. & Munchus, G. M. (1998). The predictors of quality care. Military Medicine, 163 (11), 754-757.

Walton, M. (1988). The Deming management method. New York: Putnam Publishing Group.

Ware, J. E. (1978). Effects of acquiescent response set on patient satisfaction ratings. Medical Care, 16, (4), 327-336.

Ware, J. E. & Davies, A. R. (1983). Behavioral consequences of consumer dissatisfaction with medical care. Evaluation and Program Planning, 6, 291-297.

Ware, J. E., Davies-Avery, A., & Stewart, A. L. (1978). The measure and meaning of patient satisfaction. Health and Medical Care Services Review, 1, 1-15.

Ware, J. E. & Hays, R.D. (1988). Methods for measuring patient satisfaction with specific medical encounters. Medical Care, 26, 393-402.

Ware, J. E., Orgren, R., & Conner, J. (1999). Popular patient satisfaction surveys critically reviewed. Retrieved 23 August 1999 on-line from <http://www.sf-36.com/news/sat-review.html>

Ware, J. E. & Snyder, M. K. (1975). Dimensions of patient attitudes regarding doctors and medical care services. Medical Care, 13, 669-682.

Ware, J. E., Snyder, M. K., & Wright, W. R. (1976). Development and validation of scales to measure patient satisfaction with health care services: Volume 1 of a final report. Part A. Review of literature, overview of methods and results regarding construction of scales. Carbondale, IL: Southern Illinois University.

Ware, J. E., Snyder, M. K., & Wright, W. R., & Davies, A. R. (1983). Defining and measuring patient satisfaction with medical care. Evaluation and Program Planning, 6, 247-264.

Ware, J. E., Wright, W. R., & Snyder, M. K., & Chu, G. C. (1975) Consumer perceptions of health care services: Implications for academic medicine. Journal of Medical Education, 50, 839-848.

Zapka, J. G., Palmer, R. H., Hargraves, J. L., Nerenz, D., Frazier, H. S., & Warner, C. K. (1995). Relationships of patient satisfaction with experience of system performance and health status. Journal of Ambulatory Care Management, 18 (1), 73-83.

Zimmerman, D. R., Zimmerman, P., & Lund, C. (1996). The healthcare service revolution: The growing impact of managed care on patient satisfaction. Chicago, IL: Irwin Professional Publishing.

Zyzanski, S. J., Hulka, B. S. & Cassel, J. C. (1974). Scale for the measurement of “satisfaction” with medical care: Modifications in content format and scoring. Medical Care, 12, 611-620.